Running head: ACTION PREDICTION BASED ON STATISTICAL LEARNING

Translating visual information into action predictions: Statistical learning in action and non-action contexts

Claire D. Monroy[a, c], Sarah A. Gerson[a, b] and Sabine Hunnius[a]

[a]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen

Montessorilaan 3, 6500HE Nijmegen, The Netherlands, s.hunnius@donders.ru.nl

[b]School of Psychology, Cardiff University, 70 Park Place, Cardiff, CF10 3AT, GersonS@cardiff.ac.uk

[c]Department of Otolaryngology, Ohio State University Wexner Medical Center, claire.monroy@osumc.edu

Correspondence concerning this article should be addressed to: Claire Monroy, Department of Otolaryngology, Ohio State University Wexner Medical Center, Email: claire.monroy@osumc.edu

Abstract

Humans are sensitive to the statistical regularities in action sequences carried out by others. In the current eye-tracking study, we investigated whether this sensitivity can support prediction of upcoming actions when observing unfamiliar action sequences. In two between-subjects conditions, we examined whether observers would be more sensitive to statistical regularities in sequences performed by a human agent vs. self-propelled 'ghost' events. Secondly, we investigated whether regularities are better learned when associated with contingent effects. Both implicit and explicit measures of learning were compared between agent and ghost conditions. Implicit learning was measured via predictive eye movements to upcoming actions or events, and explicit learning was measured via uninstructed reproduction of action sequences and verbal reports of the regularities. Findings revealed that participants, regardless of condition, readily learned the regularities and made correct predictive eye movements to upcoming events during online observation. However, different patterns in explicit learning outcomes emerged following observation: participants were most likely to recreate the sequence regularities and to verbally report them when they observed an actor create a contingent effect. These results suggest that the shift from implicit predictions to explicit knowledge of what has been learned is facilitated when observers perceive another agent's actions and when these actions cause effects. Findings are discussed with respect to the potential role of the motor system in modulating how statistical regularities are learned and used to modify behavior.

Keywords: action prediction, action sequences, statistical learning, implicit and explicit learning, eye-tracking

1.0 Introduction

Predicting the behavior of other people is central to social cognition and interaction. As we observe others, we automatically predict the unfolding movements and future course of their actions (Flanagan & Johansson, 2003). In everyday life, many of the actions we observe are embedded within continuous, temporal sequences. Imagine the act of baking a cake: this action is comprised of a continuous stream of individual action steps such as gathering ingredients, measuring them into bowls, mixing things together, pouring batter into a tin, and so forth. The ability to anticipate the upcoming events in a sequence is an indicator that the observer possesses some knowledge of the overarching structure of the global action and the relations between the individual steps. Perceiving the boundaries of the distinct elements in a sequence and anticipating what follows is crucial for our cognitive system to perceive the overarching activity as coherent and meaningful (Zacks & Tversky, 2001). In the current study, we investigated whether statistical regularities in novel, unfamiliar sequences support the ability to generate predictions of future events during observation[1]. Specifically, we investigated whether observers make anticipatory gaze fixations to upcoming action events based on their transitional probabilities alone, and whether they recreate learned regularities in their own action performance following observation.

1.1 Statistical learning in the domain of action

Statistical learning (SL) refers to the ability to detect regularities from structured input and operates across sensory domains (Conway & Christiansen, 2005; Krogh, Vlach, & Johnson, 2013). From early in life, humans are sensitive to multiple sources of statistical information in visual and auditory stimuli (Saffran, Johnson, Aslin, & Newport, 1999). Converging evidence indicates that SL skills are rapid and automatic, often occurring without the learner being consciously aware that he or she has learned anything at all (Turk-Browne, Scholl, Chun, & Johnson, 2008). This has led to the assumption that SL is a domain-general mechanism, with similar underlying computations and outcomes across sensory modalities. However, there is also evidence that the outcomes of SL are specific to the modality in which the stimuli are learned. For instance, one study (Conway & Christiansen, 2006) presented participants with auditory, tactile, and visual sequences defined by respective artificial grammars. Findings showed that sensitivity to statistical features was specific to each sensory modality, suggesting that SL involves "distributed, modality-constrained subsystems" (Conway & Christiansen, 2006; p.911).

Does sensitivity to statistical regularities extend to the domain of action? If so, does SL operate in a domain-general manner across all forms of perceptual events, or are there specialized subsystems that might facilitate SL particularly for observed *actions*? An initial study on action sequence processing by Baldwin and colleagues (2008) demonstrated that

[1] Unlike the cake example above, the sequences used in the current study were abstract in the sense that they did not lead to a global action goal. This was to ensure that predictions could only be based on acquiring knowledge of the sequence regularities rather than prior knowledge about the overarching event structure.

observers can rely on statistical regularities to segment action streams into discrete steps, even when transitional probabilities are the only information available for identifying action segments. At a group level, participants' performance on this action segmentation task was comparable with performance on similar tasks in the language domain. Developmental research has demonstrated similar findings with preverbal infants (Roseberry, Richie, Hirsh-Pasek, Golinkoff, & Shipley, 2011; Saylor, Baldwin, Baird, & LaBounty, 2007; Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014), showing that these segmentation skills emerge early in development. Similarity in performance across studies has led researchers to speculate that a common "statistical tracking mechanism" (Baldwin, Andersson, Saffran, & Meyer, 2008, p. 1404) is shared between processing of action and processing of other forms of perceptual stimuli.

Segmentation reveals whether observers demonstrate sensitivity to the sequence structure after learning has occurred. Typical paradigms measure segmentation by the ability to remember the items they had observed during a previous learning phase (e.g., Baldwin et al., 2008; Saffran et al., 1997). However, current theories of action perception claim that continual, automatic prediction of upcoming actions is a central feature of action processing (Kilner, Friston, & Frith, 2007a, 2007b). Importantly, predicting the outcomes of ongoing actions requires integrating prior knowledge about the most likely outcomes of the action with incoming perceptual input. Though active motor experiences are one important source of action knowledge (Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006; Libertus & Needham, 2010; Sommerville, Woodward, & Needham, 2005), motor experience alone is insufficient to explain the full range of infant and adults' capabilities for learning about actions (Hunnius & Bekkering, 2014). Statistical learning skills are therefore a candidate mechanism for how humans learn and generate predictions about upcoming action steps when observing novel, unfamiliar sequences (Ahlheim, Stadler, & Schubotz, 2014), though direct evidence for this does not yet exist. As we discuss below, we hypothesized that observing human action engages specialized cognitive processes that particularly facilitate learning of observed action sequences, relative to visual event sequences.

1.2 Outcomes of learning: implicit and explicit measures

The outcomes of SL have long been a topic of debate; in particular, discussions focus on whether and under what conditions SL results in explicit or implicit learning outcomes (Perruchet & Pacton, 2006). Typical findings have shown that SL usually occurs automatically and without conscious intent; people are often unaware of the regularities they have learned (e.g., Haider et al., 2014; Turk-Browne, Jungé, & Scholl, 2005; Turk-Browne et al., 2008). Behavioral indicators of implicit learning are typically revealed in faster reaction times (Fiser & Aslin, 2002) or anticipatory eye movements (Marcus, Karatekin, & Markiewicz, 2006) and participants are usually unaware of the subtle changes in their own behavior as a result of learning. On the other hand, SL can also result in explicit knowledge about what was learned (Bertels, Franco, & Destrebecqz, 2012; Esser & Haider, 2017b). Explicit learning is typically measured by recognition or recall which requires "conscious, or deliberate, access to memory for previous experiences" (Gomez, 1997, p. 166). In the current study, we assessed multiple measures of learning to explore how the learned information is transferred into behavior. If participants learned the statistical regularities, they could in principle predict what would occur next and shift their gaze to the next event in the sequence. If implicit knowledge from observation can be accessed and used to modify behavior, participants could also reproduce the observed regularities and report knowledge about the sequence structure.

1.3 The role of the motor system during action observation

Observing actions engages neural networks that differ from those involved in general visual and attention processes (Adams, Shipp, & Friston, 2013; Ahlheim et al., 2014;

Schubotz & von Cramon, 2009). For instance, neuroimaging research has revealed the existence of a network of sensorimotor brain regions, collectively termed ‘action-observation network’ (AON), which are specifically engaged when observing another person’s actions (Gallese & Goldman, 1998; Kilner, 2011). Activity in the AON, also sometimes termed ‘*motor resonance*’ (Rizzolatti & Craighero, 2004) or ‘simulation’ (Blakemore & Decety, 2001), is thought to facilitate prediction of observed actions by simulating how one would perform the action oneself. Predictive accounts of the motor system propose that we employ our own motor system using an internal, feed-forward model to predict the behavior of other people we observe (e.g., Kilner et al., 2007b).

In the context of embodied accounts of action observation, the motor system facilitates efficient transformation of visual information into action knowledge in the observer’s motor system. Supporting evidence from a separate line of research on observational learning shows that observers are consistently better at imitating and learning novel tool functions when observing a human actor relative to any other form of visual observation (for a review, see Hopper, 2010). These behavioral studies employed the use of a so-called ‘ghost display’, a method in which objects appear to move on their own with no agent intervention. In the current study, we adopted the ghost-display method to test the hypothesis that the learning advantage when observing another human, relative to a non-agent ghost display, extends to action predictions based on statistical learning.

1.4 The role of effects in continuous action sequences

Goal-directed actions typically result in perceivable effects, such as the sound of a whistle as it is blown. Through repeated observation, these effects become linked to the actions that consistently precede them and create ‘bidirectional action-effect associations’ (Elsner & Hommel, 2001). Prior research suggests that it is the effects of actions themselves that people anticipate when planning their own movements (Hommel, 1996). In the field of implicit learning research, action-effects have been shown to enhance implicit sequence learning when participants own motor responses result in predictable action-effects (e.g., Haider, Eberhardt, Esser, & Rose, 2014). Recent work suggests that they may also be particularly important for transferring learning from implicit into explicit awareness (Esser & Haider, 2017a, 2017b). These findings demonstrate that action-effect associations likely play a central role in establishing the contextual knowledge needed for making action predictions. Though much of this work has investigated action-effects in sequence learning of motor responses (e.g., using the standard *serial reaction time task*), there is also evidence to suggest that action-effects also guide our predictions during observation alone (Paulus, van Dam, Hunnius, Lindemann, & Bekkering, 2011).

How do sensory effects influence *observers'* sensitivity to statistical regularities when they are embedded within continuous sequences, as is the case during daily real-life perception? Based on ideomotor theory (James, 1890) and the related action-effect principle (Hommel, 1996), observers should be better at learning action contingencies when they are paired with an effect even when they do not produce the effects themselves. A matter that has not received much attention, however, is the fact that non-action visual events also result in sensory effects, such as a crashing wave. So far, we have defined effects as *action*-effects to be consistent with prior research, but it is possible that sensory effects lead to similar bidirectional associations in any form of perceptual sequence. In fact, another recent theory (Schubotz, 2007) suggests that prediction of sensory effects occurs within our sensorimotor system and can be generalized to any form of perceptual event, whether action or not. On the other hand, as we described above, evidence for enhanced learning from observing action suggests action-effects should be perceived and learned qualitatively differently than the effects of non-action perceptual events. In the current study, we manipulated whether statistical regularities were paired with an action-effect to investigate the importance of observed effects for action predictions.

1.5 The current experiment

The central focus of this study was to investigate whether observers spontaneously exploit statistical information in continuous action sequences to predict upcoming actions. Our experiment included two manipulations in order to target two primary components of action processing: (a) the role of observing an actors versus a ghost display (Agent and Ghost conditions; between-subjects), and (b) the influence of action effects versus lack of effects (Effect and No-effect pairs; within-subjects). These were assessed using an anticipatory fixation eye-tracking paradigm during action observation, which has been established as a measure of visual predictions (Hunnius & Bekkering, 2010). In addition, we examined the link between predictive looking during observation and subsequent action production. For this third aim, post-observation action performance and verbal reports were analyzed as complementary measures of implicit and explicit learning.

2.0 Method

2.1 Participants

Fifty university students participated in this study (25 in each condition [Agent and Ghost]; 43 females; *M* = 20.07 years, *range* = 18-25 years, *SD* = 2.29). Participants were recruited via an online system for students at the university and were awarded course credit for participation. Seven participants were excluded from analyses for not meeting the inclusion requirements for total looking time (see *Analysis* section), resulting in 43 participants in the final sample (23 in the Agent condition and 20 in the Ghost condition).

2.2 Stimuli

Participants' eye movements were recorded with a Tobii T60 eye-tracker (Tobii, Stockholm, Sweden) with a 17" monitor. Participants sat approximately 60cm away from the screen. Stimuli were presented with Tobii ClearView AVI presentation software and sounds were played through external speakers.

Participants observed a full-screen (1280x1024 pixels) film of a sequence involving a multi-object device that afforded six unique manipulations and a central, star-shaped light (Figure 1). To avoid confusion, we will subsequently refer to the individual object manipulations in the sequence as ‘events’, as in one condition they were human actions and in the other they were object movements. The movies were filmed with a Sony HandyCam video camera and edited using Adobe Premiere Pro Cs5 software. The same device used during filming was presented to participants before and after the observation phase.

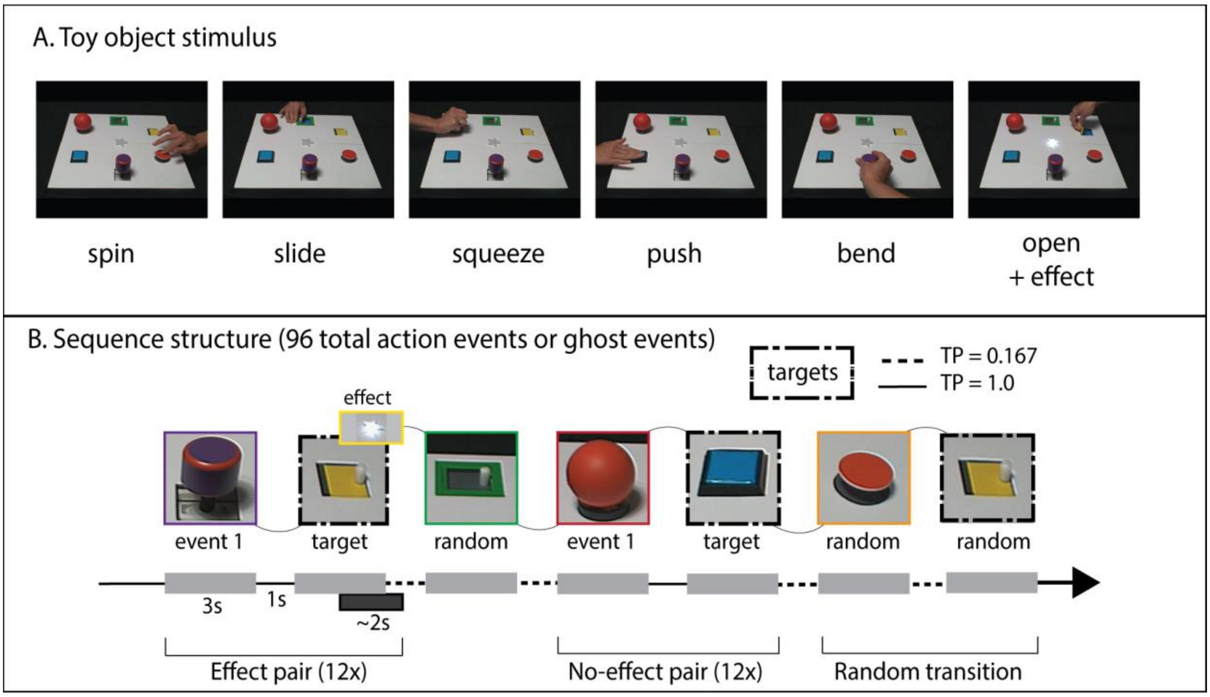


Figure 1. Overview of the experimental design.
A: Example frames from the video stimuli of the Agent condition. B: Schematic illustrating the deterministic pairs and transitional probabilities within sequences during the observation phase.

2.2.1 Sequence

We constructed four pseudo-randomized sequences, using the program *Mix* (van Casteren & Davis, 2006). All sequences contained two deterministic pairs (transitional probability between events = 1.0), labelled ‘Effect’ and ‘No-effect’ pairs (described in more detail in the following paragraph). The second event of each deterministic pair was labelled a *target*, as these were the events that became predictable as the sequence unfolded. All other possible random pairs occurred with equal frequency (transitional probabilities between events = 0.167; Figure 1B). No event or pair could occur more than three times consecutively. All pairs and random events occurred 12 times (targets thus occurred 12 times within pairs and 12 times outside of pairs). In total, participants viewed 24 deterministic pairs (12 Effect and 12 No-effect pairs) and 48 random unpaired events, for sequences of 96 total actions or events. Effect and No-effect pairs were composed of two actions that were randomly selected from the 6 possible actions. Two sets of the four sequences were created: the two actions comprising the Effect pair in one set became the No-effect pair in the second set, and vice-versa. Thus, there were eight possible sequences within each condition and 16 videos in total; participants were randomly assigned to view one of these videos.

The ‘Effect pair’ caused a central star to light up, whereas the ‘No-effect pair’ caused no additional effect. We will subsequently refer to the second events of both pairs as *targets*, as these were the events that became predictable as the sequence unfolded. The effect onset occurred at a natural mid-point of the target event during the Effect pair: for example, during the target *open*, the light turned on the moment the yellow door was fully open and turned off again after it closed (see Figure 1A).

Targets could also occur elsewhere in the sequence outside of the deterministic pair (see Figure 1B). In these instances, the effect never occurred. This ensured that the second event did not independently predict the effect, and observers were required to learn the two-step pair structure to accurately predict the effect.

Each video sequence was divided into four blocks, with the viewing angle oriented from a different side of the box for each block. This was to dissociate the events (and their corresponding objects) with their spatial location, and thus ensure that the observer could not predict the next event based on its location on the screen. Each block lasted approximately 90s and consisted of 24 events. Brief cartoon animations were presented between blocks in order to reengage the participant's attention. At the beginning of a block, one 4s still frame of the stimulus was presented to allow observers to reorient to the new perspective. Movies were approximately seven minutes long. Engaging, upbeat music was played throughout the entire demonstration that did not correspond in any way to the unfolding sequence.

2.2.2 Agent condition

In the Agent condition movies, a hand manipulated the stimulus objects in a continuous sequence. For each action, the hand entered the screen closest to the object on which it acted. Each action was exactly three seconds in duration with a one-second pause between actions during which the hand was off-screen and only the stimulus was visible.

2.2.3 Ghost condition

In the Ghost condition, the objects appeared to move on their own with a spotlight focused on the current event (see Figure 2). The spotlight gradually illuminated each object just prior to its movement onset and faded again after the object ceased moving. Between ghost events, there was a 1s pause during which it was ambiguous where the spotlight would next begin to appear, which matched the period of time the actor's hand was off-screen in the Agent condition. Like the actor's hand, the spotlight cued which object would subsequently move. The intensity and focus of the spotlight was equal for all objects. The sequence order and timing of events were otherwise identical to the videos in the Agent condition.

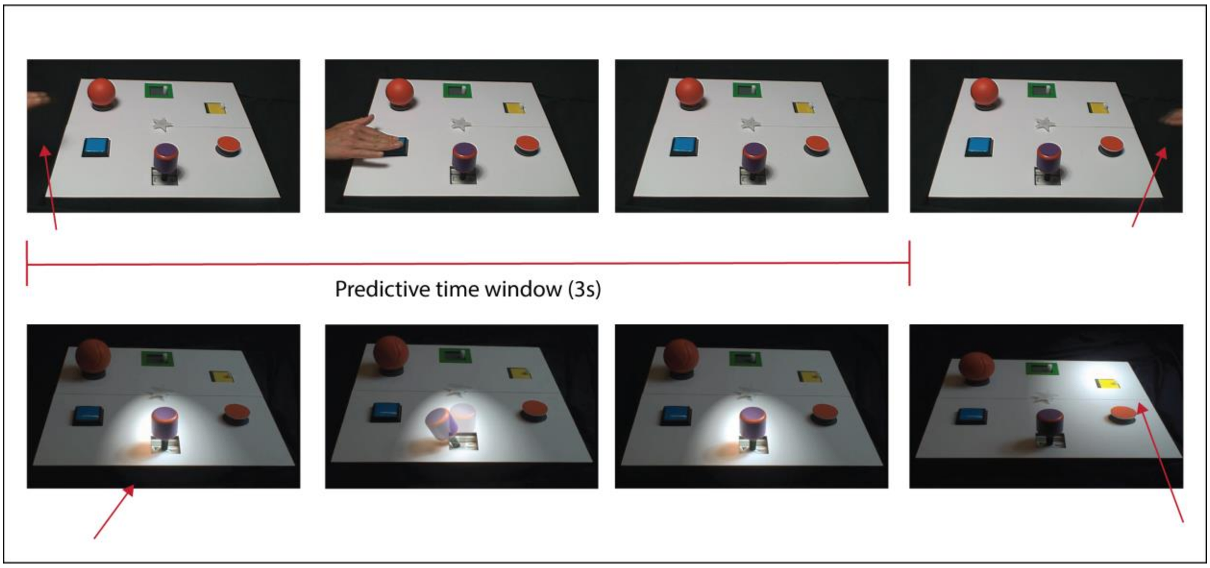


Figure 2. Predictive time windows in the Agent and Ghost conditions.
Example frames illustrating the predictive time windows in both conditions. Arrows indicate the first frame in which the agent's hand appears (Agent condition) and in which the spotlight focuses onto the target object (Ghost condition).

2.3 Procedure

Participants were first seated at a table upon which the stimulus device was placed. The side facing each participant was counterbalanced. Participants were told they would watch a video of a person interacting with the device~~,~~ and were allowed to first familiarize themselves with the objects before beginning the experiment. The side of the object facing the participant during the action execution phase was kept the same as during the initial familiarization. After familiarization, participants moved to a chair positioned in front of the eye-tracking monitor for the observation phase in which they observed the stimuli videos. First, the eye-tracker was calibrated using a standard 9-point calibration sequence provided by Tobii Studio software. Calibration was repeated until valid calibration data was acquired for at least eight calibration points. Following calibration, participants were shown one of the eight stimulus sequences. They were told that they would be shown a video but were not given specific viewing instructions.

Immediately after the observation phase, participants returned to the table and were told that they could freely interact with the stimulus for one minute (this duration was based on pilot testing). Participants were given no instruction, as our aim was to investigate whether they would spontaneously integrate observed regularities into their own actions in the absence of any task demand. The experimenter sat opposite the participant and monitored their behavior, pressing a hidden button that activated the effect (i.e., central star light) whenever he or she performed the Effect pair. After one minute, the experimenter ended the action execution phase and then asked each participant the following questions: “Do you know how to make the light turn on?” and “Did you notice any other pattern in the movies?” If participants responded “yes” they were then asked to demonstrate the correct sequence on the device. A camera facing the participant recorded this session and behavior was later coded offline to assess action performance. Each participant completed one action sequence.

3.0 Data Analysis

3.1 Eye-tracking data

Participants with total fixation time more than one standard deviation below the mean were excluded due to relative inattention to the movies. These participants yielded gaze data for less than 25% of the demonstration, which corresponded to only 3 observations of each pair and was insufficient to assess learning over the course of the experiment. This resulted in the exclusion of two participants in the Agent condition and five participants in the Ghost condition (see *Participants* section above).

Eye movement data was exported from Tobii ClearView analysis software and separated into discrete fixations using a customized software program with a spatial filter of 30 pixels and a temporal filter of 100ms. Fixation data was imported into Matlab for further analysis. Regions of interest (ROI) of identical size were defined around each object

(250x250 square pixels), and a smaller ROI (130x130 square pixels) was defined around the light (due to its smaller size relative to the objects).

For the Agent condition, fixations were considered predictive if they occurred in the time window from when the actor's hand entered the screen to perform the first action of a pair until the frame before it reappeared for the target action (Figure 2). This corresponds to the time in which the participant had enough information to predict the next action before its onset. For the Ghost condition, this time window was defined from the moment the spotlight highlighted the first object until the frame before the light shifted towards the second object of a pair. Time windows were identical in length in both conditions. As the main aim of this study was to examine prediction, only predictive gaze fixations were included in our analyses (i.e., we did not examine reactive fixations).

To assess predictive gaze during observation, we compared proportions of fixations to correct vs. incorrect objects (Implicit learning measure I). Implicit learning measure I reflects the extent to which observers predict the correct location of an upcoming event, relative to other locations. Second, we analyzed proportions of correct predictive fixations over the course of the experiment to examine how learning unfolded over time. Third, proportions of predictive fixations to target objects were compared between deterministic and random transitions (Implicit learning measure II). Learning measure II reflects the frequency of predictive looks to the target actions during predictable relative to non-predictable trials. We describe both measures in more detail below (Figure 3).

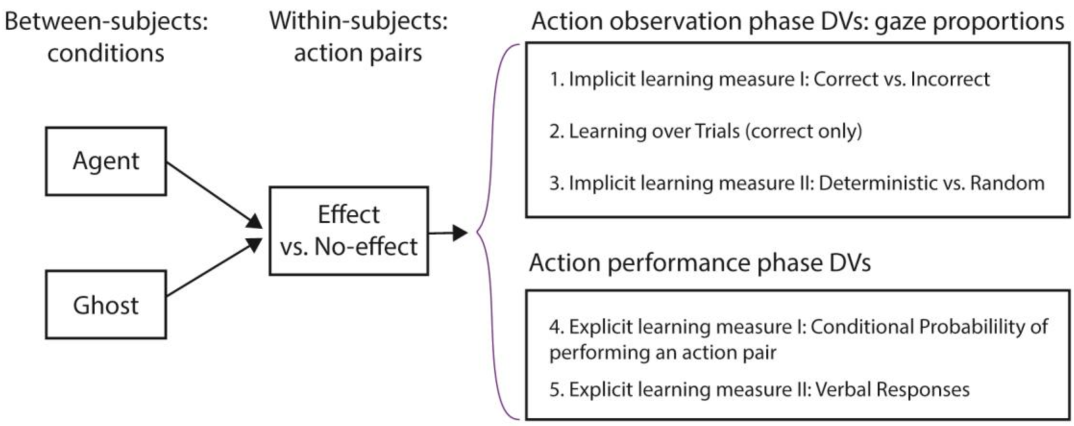


Figure 3. Overview of the experimental design and dependent variables.

3.1.1 Implicit learning measure I: Correct vs. incorrect locations

Target regions were defined around the location of the second events of each pair. Fixations to targets during predictive time windows were counted as *Correct* and fixations to the four remaining objects as *Incorrect*. Objects currently being manipulated (i.e., the first action of the pair) were excluded from analyses. The first trial of each pair was not analyzed, because participants were not expected to correctly predict the first observation of a pair. If participants learned the pair structure, we expected them to make more fixations to the locations of target objects relative to any other object during predictive time windows. For both Effect and No-effect pairs, we calculated the proportion of correct and incorrect fixations out of the total fixations to all objects (Eqs. 1 and 2). Because there were uneven numbers of correct and incorrect locations, the incorrect proportion was defined as the average number of fixations to the four remaining objects out of the total number of fixations. This location measure represents observers' bias for looking toward the correct target, relative to other objects, before it was acted upon. For additional analyses in which we included fixations to the action-effect, see Supplementary Materials.

$$Correct_{target} = \frac{\#\, fixations\ to\ target}{total\ \#\, fixations\ to\ objects} \qquad (1)$$

$$Incorrect_{target} = \frac{\#\, fixations\ to\ other\ 4\ objects/4}{total\ \#\, fixations\ to\ objects} \tag{2}$$

3.1.2 Implicit learning measure II: Deterministic vs. random transitions

Our second learning measure compared fixations to targets during deterministic vs. random trials (Eqs. 3 and 4). Random trials were defined as transitions between any possible event and the subsequent occurrence of a target event outside of a deterministic pair. We discarded all repetition trials (for example, *push* followed by *push*) because it was impossible to determine whether fixations during these trials were predictive or reactive (i.e., simply not moving the eyes). This analysis thus enabled us to compare fixations to the same location (target objects) in different statistical contexts.

$$Deterministic\text{[2]} = \frac{\#\, fixations\ to\ target\ (predictive\ trials)}{total\ \#\, fixations\ to\ objects} \tag{3}$$

$$Random = \frac{\#\, fixations\ to\ target\ (non\text{-}predictive\ trials)}{total\ \#\, fixations\ to\ objects} \tag{4}$$

3.2 Behavioral data

3.2.1 Explicit learning measure I: Action performance

Participants' self-produced action sequences were coded from the videotape recordings. Each object manipulation was counted as a single action. We calculated the conditional probability of performing the second action of a pair (*B*), given performance of the first action (*A*), to account for variation in the overall length of participants' sequences. Conditional probability was defined as:

$$P(B|A) = \frac{P(A,B)}{P(A)} \tag{5}$$

3.2.2 Explicit learning measure II: Verbal responses

Responses to the experimenters' explicit questions—"Do you know how to make the light turn on?" and "Did you notice any other pattern in the movies?"—were coded as yes or

[2] Note that this equation is identical to Eq. 1

no; if their response was yes, it was further coded as 'yes'-correct or 'yes'-incorrect depending on whether or not they demonstrated the correct sequence on the first attempt. Proportions of participants who indicated each response type were calculated for each pair, per condition.

4.0 Results

4.1 Eye movement data

To examine whether the Agent and Ghost displays elicited similar rates of overall visual attention to the objects of interest, we compared the number of predictive fixations between the two conditions. There were no differences in the number of anticipatory fixations made during target trials (Ghost = 41.55, *SEM* = 4.80; Agent: M = 44.61, *SEM* = 3.41; p = .60) or in the total number of fixations made across the entire demonstration (p = .21) suggesting that differences in the visual stimuli in the Agent and Ghost conditions did not underlie any potential differences in anticipatory fixations. Analyses of total looking times in seconds are reported in the Supplementary materials.

4.1.1 Implicit learning measure I: Correct vs. incorrect locations

Our primary learning measures in each condition are presented in Table 1. Proportions of gaze fixations were analyzed via a repeated-measure ANOVA with Prediction (Correct vs. Incorrect) and Pair (Effect vs. No-effect) as within-subject factors and Condition (Agent vs. Ghost) as a between-subjects factor. This analysis revealed a main effect of Prediction, indicating that participants made a higher proportion of correct relative to incorrect predictive fixations across pairs (mean difference = .14 [*SEM* = .04], $F(1,40)$ = 16.27, $p < .001$, $\eta_p^2 = .29$). There were no other significant main effects or interactions (*ps* > .13). The results of additional analyses including the location of the action-effect as a correct location are available in the Supplemental Information.

Table 1.

Main implicit and explicit dependent measures, separated by condition.

			Agent (N = 23)		Ghost (N = 20)	
Learning measure	Pair		Mean	*SD*	Mean	*SD*
I: Correct vs. Incorrect	Effect	Correct (Eq. 1)	0.39	*0.26*	0.34	*0.33*
		Incorrect (Eq. 2)	0.09	*0.05*	0.11	*0.07*
	No-effect	Correct (Eq. 1)	0.26	*0.28*	0.25	*0.22*
		Incorrect (Eq. 2)	0.19	*0.07*	0.19	*0.06*
II: Deterministic vs. Random	Effect	Deterministic (Eq. 3)	0.39	*0.26*	0.34	*0.33*
		Random (Eq. 4)	0.25	*0.20*	0.18	*0.16*
	No-effect	Deterministic (Eq. 3)	0.26	*0.28*	0.25	*0.22*
		Random (Eq. 4)	0.13	*0.14*	0.14	*0.12*
Action Performance	Effect	Conditional probability (Eq. 5)	0.54	*0.36*	0.30	*0.30*
	No-effect		0.29	*0.36*	0.09	*0.16*
Verbal Response (“yes” – correct)	Effect	% participants	68.4%		15.4%	
	No-effect		5.9%		7.7%	

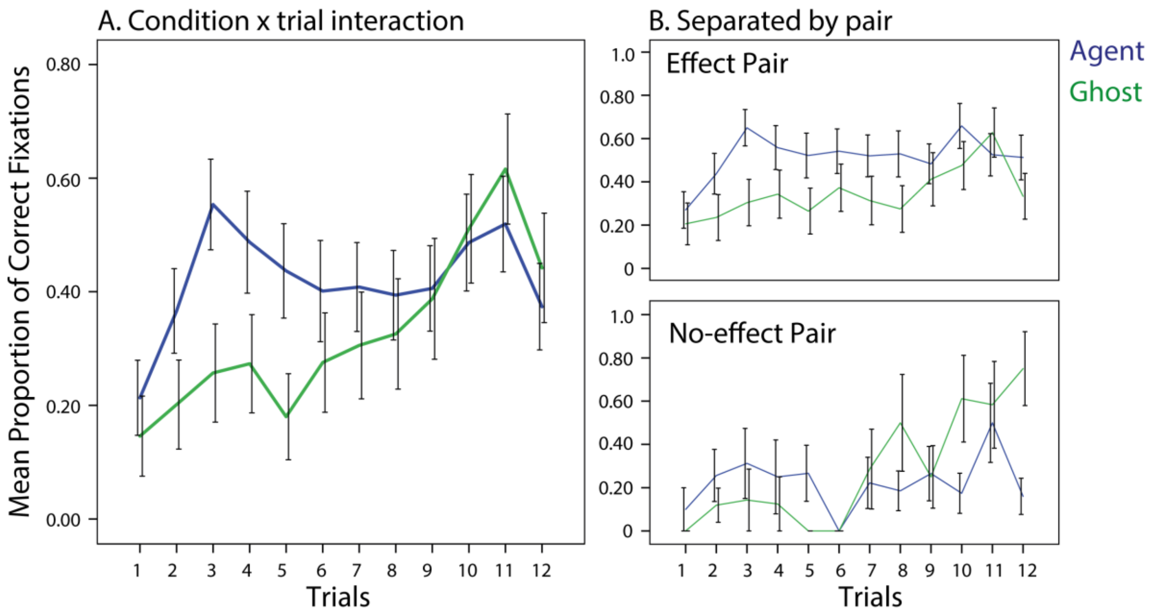
Note. For learning measure I, column 3 refers to proportions of correct and incorrect fixations. For learning measure II, column 3 refers to proportions of correct fixations on deterministic or random trials.

4.1.2 Learning over trials

To examine changes in predictions across trials, we performed a general estimating equations (GEE) analysis. GEE analyses are a preferred method for analyzing data with repeated measures that contain missing points, such as trials in which no anticipatory fixations were recorded, because they do not apply list-wise exclusion of cases (Zeger, Liang, & Albert, 1988). Proportions of correct fixations to the targets were entered as the dependent variable in a linear, model-based GEE with an unstructured Working Correlation Matrix. Condition (between-subjects), Trial (within-subjects), and Pair (within-subjects) were entered as predictors in a factorial model. In this analysis, the first trial was included (in contrast to Learning measures I and II).

The GEE analysis yielded significant main effects of Trial ($\chi 2(11) = 47.19$, $p < .001$)

and Pair ($\chi 2(2) = 26.89$, $p < .001$) a significant interaction between Condition and Trial ($\chi 2(11) = 21.52$, $p = .028$) a significant interaction between Condition and Pair ($\chi 2(2) = 8.70$, $p = .003$) and a three-way Condition by Trial by Pair interaction ($\chi 2(11) = 22.96$, $p = .02$). The Condition by Pair interaction revealed that proportions of correct fixations were significantly greater in the Agent relative to the Ghost condition for the Effect pair (mean difference = .18 [*SEM* = .05], $p < .001$) but not for the No-effect pair (mean difference = -.09 [*SEM* = .06], $p = .11$)[3]. As illustrated in Figure 4, the Condition by Trial interaction revealed that the Agent and Ghost conditions did not differ from one another on the very first ($p = .45$) or second trial ($p = .15$). By the third trial, participants in the Agent condition made more correct fixations than in the Ghost condition (mean difference = .28 [*SEM* = .12], p = .015) and this pattern continued for several trials. The two conditions converged again by the 6th trial ($p = .53$) for the remainder of the experiment. Together, these findings suggest that participants showed a selective learning benefit for making correct anticipations when viewing an agent producing action-effects, relative to the other observation contexts.



[3] Note that the interaction between Condition and Pair was not statistically significant in our first analysis (4.1.1). This is likely due to the fact that the first analysis included both correct and incorrect fixations, whereas the *Learning over Trials* analysis examines correct fixations only.

Figure 4. Learning over time.
Estimated marginal means of correct predictive fixations across pairs as a function of trial, (left) collapsed across pairs and (right) separated by Effect and No-effect pairs. Bars represent standard errors.

4.1.3 Implicit learning measure II: Deterministic vs. random transitions

The proportion of gaze fixations to target objects (Eqs. 3 and 4) were entered as the dependent variables into an ANOVA with Transition (Deterministic vs. Random) and Pair (Effect vs. No-effect) as within-subjects factors and Condition (Agent vs. Ghost) as a between-subjects factor. This revealed a main effect of Transition, showing that participants made more target fixations during deterministic than during random transitions across conditions and pairs, $F(1, 42) = 42.9$, $p < .001$, $\eta_p^2 = .51$. There were no other effects or interactions (*ps* > .11).

4.2 Explicit measures of learning

4.2.1 Explicit learning measure I: Action performance

Across conditions, participants performed sequences with an average length of 26.22 actions ($SD = 7.1$), and performed a mean of 2.12 Effect pairs and 0.64 No-effect pairs (see Table 1 for additional descriptive measures). There were no differences in the total length of action sequences performed between conditions ($p = .19$).

Conditional probabilities for performing the target action given the performance of the first action of the pair were entered in an ANOVA with Pair (Effect vs. No-effect) as a within-subjects factor and Condition (Agent vs. Ghost) as a between-subjects factor. This revealed main effects of Condition and Pair: participants in the Agent condition were more likely to perform an action pair than those in the Ghost condition, $F(1, 34) = 11.57$, $p = .002$, $\eta_p^2 = .25$ (see Figure 5a). Across conditions, participants were more likely to perform the Effect pair than the No-effect pair, $F(1, 34) = 8.25$, $p = .007$, $\eta_p^2 = .20$. There was no interaction between Pair and Condition ($p = .78$).

To assess whether participants in each group performed more pairs than would be expected by chance, we conducted a one-sample t-test to compare the mean conditional probability of performing each pair against a chance level of 0.167 (one out of six possible actions, given any previous action*)*. This revealed that the participants in the Agent condition performed Effect pairs significantly more than chance ($p < .001$), while participants in the Ghost condition did not ($p = .13$). In neither condition were the No-effect pairs performed at an above-chance level (*ps* > .05).

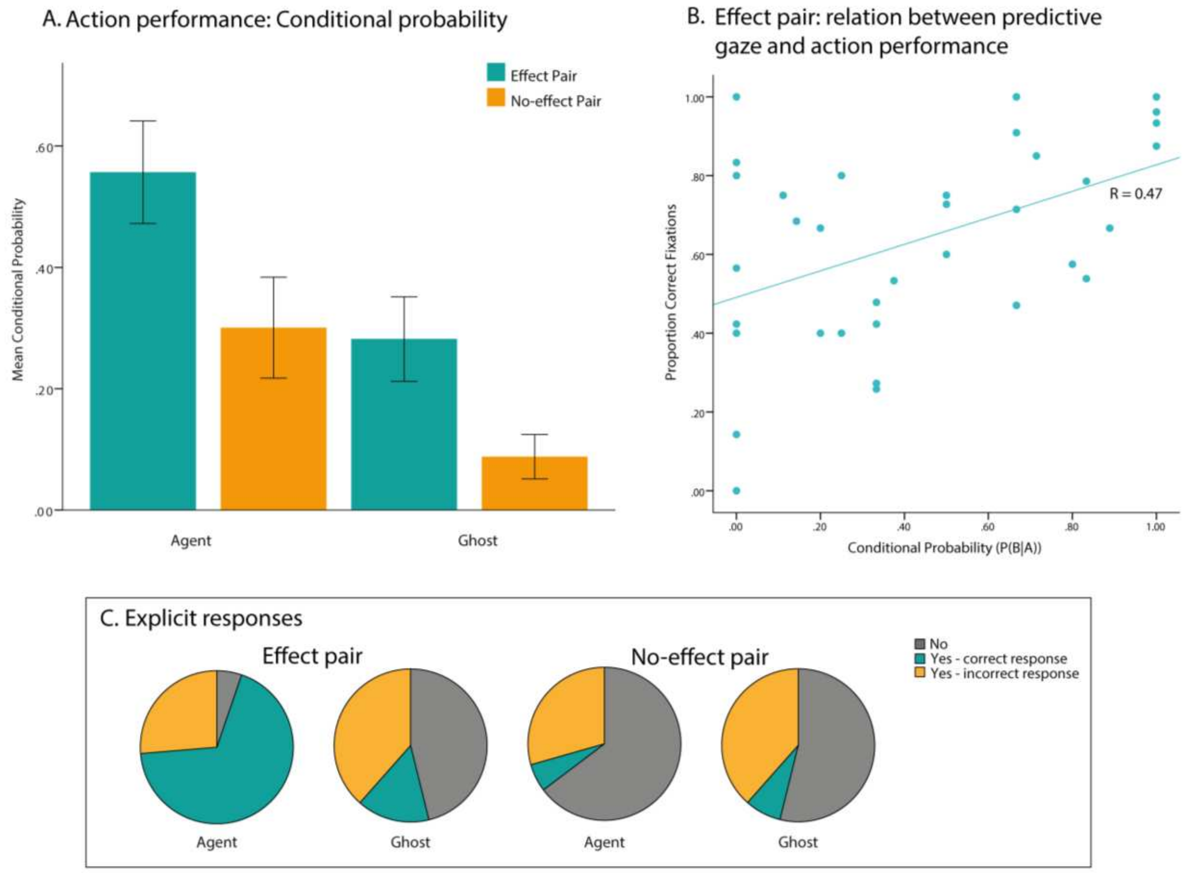


Figure 5. Action performance and verbal awareness.
A: The mean probability of performing Effect and No-effect pairs (*P(B|A)*). Bars represent standard errors. B: Scatterplot illustrating the relation between predictive fixations (Eq. 1) and action performance (Eq. 5) for the Effect pair, across conditions. C: Pie graphs showing the percentage of participants who gave each response type to the experimenter's question. For the Effect pair, this was "Do you know how the light turns on?" and for the No-effect pair this was "Did you see any other pattern in the movies?"

To investigate whether action execution was related to anticipatory looking behavior, we correlated the proportion of correct target fixations (Eq. 1) and the conditional probability of producing action pairs for each pair type. Across conditions, there was a significant positive correlation between target fixations during Effect pairs and the conditional probability of producing Effect pairs, $r(35) = .41$, $p = .02$, indicating that participants who demonstrated higher rates of learning during the observation phase were more likely to reenact the action-effect during the subsequent behavioral session (Figure 5b). There was no correlation for the No-effect pair, $r(36) = .01$, $p = .97$. These correlation coefficients differed significantly from one another, $Z = 1.75$, $p = .04$[4].

4.2.2 Explicit learning measure II: Verbal responses

Figure 5c illustrates the distributions of participants per each explicit response type to the experimenter's questions following the action execution phase, separated by pair and condition. The pie charts reflect the following pattern: 94.7% of participants in the Agent condition reported explicit knowledge of the Effect pair; of these, 72.2% were correct and 27.8% were incorrect. Only 53.8% reported explicit knowledge of the pair in the Ghost condition; 28.6% of these were correct and 71.4% were incorrect. Further, only 40% reported knowledge of the No-effect pair across conditions, and those who did were usually incorrect (93.3% of these 40%).

To compare these proportions of participants (Agent vs. Ghost) to one another, we calculated the confidence intervals of the difference between them (the difference between proportions is statistically significant wherever the confidence interval excludes zero; Newcombe, 1998; Wilson, 1927). Table 2 reports the confidence intervals for the differences in proportions for each response type. For the Effect pair, the proportion of participants who responded 'yes' and were correct was significantly greater in the Agent than the Ghost

[4] For thoroughness, we also averaged across pairs and correlated the fixation proportions with conditional probability for Agent and Ghost conditions separately. Across pairs, there were no significant correlations for either group, *p*s > .42. These correlation coefficients did not differ significantly from one another ($Z = .41$, $p = .34$).

condition. A higher proportion of participants in the Agent condition reported knowledge of the Effect pair—and could demonstrate the correct sequence—than in the Ghost condition. Likewise, significantly more participants in the Ghost condition reported *no* knowledge of the Effect pair than in the Agent condition. For the No-effect pair, the pattern of responses was similar across conditions. Thus, participants observing an actor were more likely to retain precise knowledge they could verbalize about the pair structure, but only when the actor's actions led to a causal effect. Participants observing ghost events were less likely to report verbal knowledge, and when they did, their representations of the pair structure were more likely to be inaccurate.

Table 2.

Mean differences (and confidence intervals) between conditions (Agent – Ghost) in the proportions of participants reporting each response type for Effect and No-effect pairs.

	Effect Pair		No effect Pair	
Response	Diff(P_a-P_b)	95% CI	Diff(P_a-P_b)	95% CI
"No"	-.41	**[.11, .66]***	.11	[-.22, .41]
"Yes"-correct	.53	**[.18, .73]***	-.02	[-.20, .28]
"Yes"-incorrect	-.12	[-.18, .42]	-.09	[-.22, .40]

Note. Diff(P_a-P_b) indicates the difference between the proportions of participants in the Agent and Ghost conditions. *denotes statistically significant difference between the two sample proportions (p < .05).

5.0 Discussion

The current study investigated whether observers can learn statistical regularities during observation of continuous action or event sequences. Specifically, we measured anticipatory gaze fixations as an implicit measure of whether participants could use statistical information to predict upcoming actions or events in the sequence. After learning, we measured spontaneous action performance and verbal reports as explicit measures of whether observed statistical regularities influence participants' self-produced actions and knowledge of the sequence.

5.1 Implicit learning: Predictive gaze

Across conditions and pairs, participants demonstrated a robust tendency to predict correct relative to incorrect locations. They also predicted the target more frequently during deterministic relative to random transitions between events. In other words, they looked to where a target event was statistically likely to occur next, and they looked to the targets selectively when they were likely to occur next relative to when they were unlikely to occur next.

When examining correct predictions over time, an interaction effect between these two manipulations emerged: participants appeared to learn the regularities best when they observed an actor produce an action-effect. In addition, different patterns emerged between the Agent and Ghost conditions for implicit and explicit learning outcomes, as measured by visual anticipations, action performance, and verbal knowledge of the pair structure. Specifically, observing actions in the Agent condition did not seem to uniquely benefit predictive gaze performance relative to observing visual events in the Ghost condition; however, it did increase reproduction of the action pair and verbal knowledge about the pair structure. Importantly, these differences were apparent only for the sequence pair which resulted in an action-effect. One explanation for these patterns is that action-specific processing in the Agent condition facilitated transfer from implicit (i.e., eye movements) to explicit (i.e., self-produced actions, verbal awareness) knowledge, as we discuss in the following sections.

5.2 Actions versus perceptual sequences

Participants demonstrated learning both when observing an actor and ghost events, as indicated by their correct predictive looks while observing the sequences in both conditions. This finding suggests that statistical learning operates consistently across the different types of perceptual events, both action and non-action. Interestingly, learning emerged earlier in the

Agent condition than in the Ghost condition. Consistent with prior research, this finding reveals a subtle learning benefit when observing an agent relative to other forms of visual displays (Hopper, Flynn, Wood, & Whiten, 2010; Hopper, Lambeth, Schapiro, & Whiten, 2015). According to motor-based accounts of action observation, this benefit originates from internal predictive models based in the motor system (Kilner et al., 2007; Stapel, Hunnius, Meyer, & Bekkering, 2016). Here we show that observers demonstrate faster learning in the Agent condition relative to the Event condition. Specifically, participants' rates of correct fixations to target actions increased more quickly in the Agent condition, revealing that they more easily detected the statistical relations between the actions and could modify their looking behavior accordingly. Interpreted within these motor-based accounts, this may reflect a more efficient ability to transfer knowledge acquired from visual statistical learning into action predictions that are generated in the motor system (Kilner, 2009).

As discussed in the introduction, developmental studies have shown that children learn significantly better from observing an agent performing actions relative to other forms of observational learning (Hopper, 2010). One recent study, in fact, showed that toddlers were able to learn action sequences when observing an actor, but not ghost events (Monroy, Gerson, & Hunnius, 2017). This finding may reflect an interesting developmental shift, in which actions provide a unique context that helps infants and children use acquired knowledge from statistical learning to make predictions, above and beyond other stimuli. Adults, on the other hand, are able to employ their statistical learning abilities across action and non-actions contexts. Nevertheless, observing actions seems to elicit a learning benefit that is consistent across development.

Though we made every attempt to match the stimuli in the two conditions for saliency, there could have still been perceptual differences between the Agent and Ghost conditions that could alternatively explain our findings. However, perceptual differences cannot solely explain the observed results, as we find no differences in overall visual attention or predictive fixations between conditions during observation. Secondly, both conditions demonstrated learning during observation, but those in the Agent condition specifically reproduced more action pairs and acquired more explicit sequence knowledge than participants in the Ghost condition. This finding suggests that there were qualitative differences in the way the sequence information was learned in Agent condition that are unlikely to be a result of perceptual saliency.

5.3 The role of effects

Observing an agent produce causal effects led to higher rates of verbal knowledge and reproduction of the action pair, relative to observing the ghost events or the pairs with no effect (both action and ghost). This pattern supports the interpretation that observing actions primarily influences the way in which learned knowledge is subsequently used to modify behavior. Even though participants were uninstructed, observing an actor produce an effect in the world may have automatically induced participants to perceive these events as goal-directed, and to attempt to re-create them in the test setting. An alternative explanation, suggestive of lower-level accounts, is that the action-effect simply provides additional information and is therefore easier to learn. The action-effect relation contains more information (i.e., A predicts both B and C) than the action-only pair (A predicts B). In addition, the action-effect contingency contains an additional dimension (i.e., actions and effects versus only actions). According to the model of sequence learning given by Keele and colleagues (2003), multidimensional learning requires additional attention components that are not required during unidimensional learning. These attentional requirements enhance sequence learning by making the learned information accessible to explicit awareness (Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003).

When only analyzing correct predictions over time, an interaction effect emerged which revealed that participants in the Agent condition demonstrated more correct predictive fixations for the Effect relative to the No-effect pair, whereas this pattern did not hold for the participants in the Ghost condition. However, this interaction effect did not appear when comparing fixations to both correct and incorrect locations. One possible explanation for this inconsistency is that, in the absence of a visual effect, participants were free to engage in more visual exploratory behaviors to the other objects, resulting in higher proportions of incorrect fixations for the No-effect pair relative to the Effect pair.

5.4 Action performance and its relation to prediction

Across conditions, participants were more likely to reproduce the pair associated with an effect than the pair without an effect. In addition, rates of performing the effect pair were correlated with participants' predictive looking for this pair. Specifically, the more accurately observers predicted the Effect pair, the more likely they were to reproduce the effect following observation. Adults and children easily recreate effects that they see in the world when explicitly asked to do so; this has been empirically demonstrated in both forced-choice and free-choice designs for simple action-effect contingencies (Elsner, 2007; Elsner & Hommel, 2001). Here, our results provide new evidence that observers could recreate action-effects based only on learning transitional probabilities, and they did so in the absence of instruction or any explicit task. These findings suggest that new action knowledge—acquired via observational statistical learning—can be accessed and used for action control when the learned actions are used for produced a desired effect or outcome.

In addition, participants in the Agent condition were more likely to reproduce action pairs than participants in the Ghost condition. This was not due to a general difference in activity between the two conditions, as they did not simply perform more actions overall. Based on the idea that we naturally tend to perceive human behavior as goal-directed, observers in the Agent condition may have automatically attributed meaning to the actor's actions and were more motivated to imitate what they observed, especially when they resulted in an effect (Hopper, 2010; Hopper et al., 2014). Alternatively, consistent with the faster emergence of correct anticipations in the Agent condition, these participants may have also been better able to retain the new knowledge gained from the observed sequence and apply it when performing their own action sequences than those in the Ghost condition.

5.5 Relations between predictive gaze, action performance, and verbal knowledge

Whether statistical learning engages implicit or explicit processes—and whether the resulting knowledge is also implicit or explicit—is an ongoing debate (see Daltrozzo & Conway, 2014 for a review). In the current study, we measured predictive gaze, action performance, and verbal responses as reflecting different learning outcomes. These behaviors may also relate to varying levels of implicit and explicit knowledge of the learned structure. Studies on SL typically demonstrate that the outcomes of learning, and thus the learning processes, are manifested in implicit behaviors such as anticipatory gaze, if at all (Fiser & Aslin, 2001; Perruchet & Pacton, 2006; Turk-Browne et al., 2008). Currently, there is a divide between those who argue that SL is an implicit mechanism (e.g., Clegg, DiGirolamo, & Keele, 1998) and those who suggest that the process may be implicit but the knowledge obtained via SL can become explicit when, for instance, learning reaches a certain threshold (Cleeremans, 2006). In the former case, it is argued that knowledge can only become explicit when other cognitive systems come into play. Recent findings have shown that sequence learning also results in explicit knowledge depending on the 'task set'; that is, the relation between the stimulus characteristics and the required response of the learner (Esser & Haider, 2017a, 2017b).

Consistent with these recent findings, our data suggest that observing action sequences results in both implicit and explicit learning outcomes[5]. One possibility, grounded in predictive accounts of the motor system, is that the knowledge gained via statistical learning can be accessed by the motor system and used to update internal action models. These models serve to generate predictions about the most likely upcoming action and to prepare appropriate motor responses. Our findings differ from prior research in that, in the current experiment, no response was required from participants during observation. Thus, the resulting explicit knowledge did not arise from learned stimulus-response associations (as in Haider et al., 2014). Rather, observation alone was sufficient to elicit both implicit and explicit knowledge. Further, our findings suggest that observing human actions facilitates both implicit sequence learning (indicated by faster learning rates in the Agent condition) and transferring learned knowledge into explicit responses. However, as suggested by Schubotz (2007), motor-based learning and prediction can still occur for external events (i.e., non-actions). A fascinating question for further research is whether observing action sequences engages entirely distinct learning processes from other forms of observational learning, or whether the difference mainly lies in how the knowledge is accessed and used. Another possibility to be considered is that acting immediately prior to being questioned by the experimenter may have influenced some participants' verbal knowledge. That is, action performance may have helped them to verbalize knowledge that otherwise would have remained implicit. However, if there was an effect of acting on participants' explicit knowledge of the sequence, this should have been consistent across conditions. Instead, the dramatic group differences in verbal knowledge that we observed suggest that responses were

[5]As we did not directly measure the learning processes, but rather the learning outcomes, we cannot speak to whether or not the learning processes themselves were implicit or explicit and focus our discussion on the outcomes of learning.

primarily influenced by the action observation condition, rather than their own action production.

5.6 Conclusion

The current study investigated whether SL abilities can support online prediction during action observation. In particular, we compared observers' sensitivity to statistical regularities in action sequences when observing a human actor relative to visual events. Our main finding revealed that implicit learning occurred in both observation conditions and was not dependent on action-effects; however, explicit knowledge was only consistently extracted when observers viewed a human actor perform action sequences with causal effects. These findings shed light on the potential role of the motor system in enhancing how information learned solely via observation can be accessed and used to modify behavior.

Acknowledgments

This work was supported by an Initial Training Network (ITN) of the People Marie Curie Actions—Seventh Research Programme (FP7) of the European Union (FP7ITN2011-289404). We thank the undergraduates of Radboud University who participated in the current study, as well as members of the BabyBRAIN research group for feedback on data analyses.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Ahlheim, C., Stadler, W., & Schubotz, R. (2014). Dissociating dynamic probability and predictability in observed actions-an fMRI study. *Frontiers in Human Neuroscience*, *8*, 273. https://doi.org/10.3389/fnhum.2014.00273

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407. https://doi.org/10.1016/j.cognition.2007.07.005

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1425–1431. https://doi.org/10.1037/a0027210

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*(8), 561–567. https://doi.org/10.1038/35086023

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(19), 1905–1910. https://doi.org/10.1016/j.cub.2006.07.065

Cleeremans, A. (2006). Conscious and unconscious cognition: A graded, dynamic perspective. *International Journal of Psychology*, *1*, 401–418. https://doi.org/10.4324/9780203783122

Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in Cognitive Sciences*, *2*(8), 275–281. https://doi.org/10.1016/S1364-6613(98)01202-9

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Frontiers in Human Neuroscience*, *8*, 437. https://doi.org/10.3389/fnhum.2014.00437

Elsner, B. (2007). Infants' imitation of goal-directed actions: The role of movements and action effects. *Acta Psychologica*, *124*(1), 44–59. https://doi.org/10.1016/j.actpsy.2006.09.006

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 229–240. https://doi.org/doi.org/10.1037/0096-1523.27.1.229

Esser, S., & Haider, H. (2017a). Action–effects enhance explicit sequential learning. *Psychological Research*, 1–17. https://doi.org/10.1007/s00426-017-0883-5

Esser, S., & Haider, H. (2017b). The emergence of explicit knowledge in a serial reaction time task: The role of experienced fluency and strength of representation. *Frontiers in Psychology*, *8*, 502. https://doi.org/10.3389/fpsyg.2017.00502

Fiser, J., & Aslin, R. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

Fiser, J., & Aslin, R. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gomez, R. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, *33*(2), 154–207. https://doi.org/10.1006/cogp.1997.0654

Haider, H., Eberhardt, K., Esser, S., & Rose, M. (2014). Implicit visual learning: How the task set modulates learning by determining the stimulus-response binding. *Consciousness and Cognition*, *26*(1), 145–161. https://doi.org/10.1016/j.concog.2014.03.005

Hommel, B. (1996). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological Research*, *59*(3), 176–186. https://doi.org/10.1007/BF00425832

Hopper, L. M. (2010). “Ghost” experiments and the dissection of social learning in humans and animals. *Biological Reviews*. https://doi.org/10.1111/j.1469-185X.2010.00120.x

Hopper, L. M., Flynn, E. G., Wood, L., & Whiten, A. (2010). Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. *Journal of Experimental Child Psychology*, *106*(1), 82–97. https://doi.org/10.1016/j.jecp.2009.12.001

Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2015). The importance of witnessed agency in chimpanzee social learning of tool use. *Behavioural Processes*, *112*, 120–129. https://doi.org/10.1016/j.beproc.2014.10.009

James, W. (1890). The Principles of Psychology. *Journal of the History of Philosophy*, *21*(2), 270–272. https://doi.org/10.1353/hph.1983.0040

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, *110*(2), 316–339.

https://doi.org/10.1037/0033-295X.110.2.316

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–357. https://doi.org/10.1016/j.tics.2011.06.005

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). The mirror-neuron system: a Bayesian perspective. *Neuroreport*, *18*(6), 1–5. https://doi.org/10.1097/WNR.0b013e3281139ed0

Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, *3*, 598. https://doi.org/10.3389/fpsyg.2012.00598

Libertus, K., & Needham, A. (2010). Teach to reach: The effects of active vs. passive reaching experiences on action and perception. *Vision Research*, *50*(24), 2750–2757. https://doi.org/10.1016/j.visres.2010.09.001

Marcus, D. J., Karatekin, C., & Markiewicz, S. (2006). Oculomotor evidence of sequence learning on the serial reaction time task. *Memory & Cognition*, *34*(2), 420–432. https://doi.org/10.3758/BF03193419

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Statistics in Medicine*, *17*(8), 873–890.

Paulus, M., van Dam, W., Hunnius, S., Lindemann, O., & Bekkering, H. (2011). Action-effect binding by observational learning. *Psychonomic Bulletin & Review*, *18*(5), 1022–1028. https://doi.org/10.3758/s13423-011-0136-3

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Science*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Roseberry, S., Richie, R., Hirsh-Pasek, K., Golinkoff, R. M., & Shipley, T. F. (2011). Babies catch a break: 7- to 9-month-olds track statistical probabilities in continuous dynamic events. *Psychological Science*, *22*(11), 1422–1424. https://doi.org/10.1177/0956797611422074

Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., Aslin, R., Tunick, R. a., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Saylor, M., Baldwin, D., Baird, J., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition and Development*, *8*(1), 113–128. https://doi.org/10.1080/15248370709336996

Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, *11*(5), 211–218. https://doi.org/10.1016/j.tics.2007.02.006

Schubotz, R. I., & von Cramon, D. Y. (2009). The case of pretense: Observing actions and inferring goals. *Journal of Cognitive Neuroscience*, *21*(4), 642–653. https://doi.org/10.1162/jocn.2009.21049

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1). https://doi.org/10.1016/j.cognition.2004.07.004

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014).

Infants segment continuous events using transitional probabilities. *Child Development*, *85*(5), 1821–1826. https://doi.org/10.1111/cdev.12247

Stapel, J. C., Hunnius, S., Meyer, M., & Bekkering, H. (2016). Motor system contribution to action prediction: Temporal accuracy depends on motor experience. *Cognition*, *148*, 71–78. https://doi.org/10.1016/j.cognition.2015.12.007

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology*, *134*(4), 552–64. https://doi.org/10.1037/0096-3445.134.4.552

Turk-browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2008). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 1934–1945. https://doi.org/10.1162/jocn.2009.21131

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, *44*(4), 1049–1060. https://doi.org/10.2307/2531734

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.